



# Weak disjointness of measure preserving dynamical systems

Emmanuel Lesigne, Benoît Rittaud, Thierry de La Rue

## ► To cite this version:

Emmanuel Lesigne, Benoît Rittaud, Thierry de La Rue. Weak disjointness of measure preserving dynamical systems. *Ergodic Theory and Dynamical Systems*, 2003, 23. hal-00017086

**HAL Id: hal-00017086**

**<https://hal.science/hal-00017086>**

Submitted on 16 Jan 2006

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# WEAK DISJOINTNESS OF MEASURE PRESERVING DYNAMICAL SYSTEMS

E. LESIGNE, B. RITTAUD, AND T. DE LA RUE.

ABSTRACT. Two measure preserving dynamical systems are weakly disjoint if some pointwise convergence property is satisfied by ergodic averages on their direct product (a precise definition is given below). Disjointness implies weak disjointness. We start studying this new concept, both by stating some general properties and by giving various examples. The content of the article is summarized in the introduction.

## CONTENTS

1. Introduction	2
1.1. Definition	2
1.2. Motivation	2
1.3. Brief description of the content	2
1.4. Questions	3
1.5. Acknowledgments	3
2. Maximal inequality, disjointness and weak disjointness	4
3. Common factors and relative disjointness	6
3.1. Some facts about factors and joinings	6
3.2. $T$ -factors and relative disjointness	8
3.3. $T$ -factors and weak disjointness	9
4. Examples of weak disjointness	11
4.1. Quasi-discrete spectrum dynamical systems	11
4.2. Chacon's transformation	12
4.3. Unipotent transformations	14
4.4. The dynamical system associated to the Morse sequence	14
5. Examples of lack of weak disjointness	15
5.1. Positive entropy	15
5.2. A cutting and stacking procedure	16
5.3. A skew-product construction	18
5.4. Rank one constructions	19
References	25

## 1. INTRODUCTION

**1.1. Definition.** In this article we call a *dynamical system* any probability measure preserving dynamical system on a Lebesgue space : a dynamical system is a quadruple  $(X, \mathcal{A}, \mu, T)$  where  $(X, \mathcal{A}, \mu)$  is a Lebesgue probability space and  $T$  is a measurable transformation of  $(X, \mathcal{A})$  which preserves the measure  $\mu$ . When there will be no ambiguity, this dynamical system will be denoted by the symbol  $T$  alone.

**Definition 1.** Two dynamical systems  $(X, \mathcal{A}, \mu, T)$  and  $(Y, \mathcal{B}, \nu, S)$  are *weakly disjoint* if, given any function  $f$  in  $L^2(\mu)$  and any function  $g$  in  $L^2(\nu)$ , there exist a set  $A$  in  $\mathcal{A}$  and a set  $B$  in  $\mathcal{B}$  such that

$$(1.1) \quad \begin{aligned} &\bullet \mu(A) = \nu(B) = 1 \\ &\bullet \text{for all } x \in A \text{ and for all } y \in B, \text{ the sequence} \\ &\quad \left( \frac{1}{N} \sum_{n=0}^{N-1} f(T^n x) \cdot g(S^n y) \right)_{N>0} \end{aligned}$$

converges.

Note that, by Birkhoff's ergodic theorem applied to the cartesian product of the dynamical systems  $T \times S$ , we know that for  $\mu \otimes \nu$ -almost all  $(x, y)$  the sequence (1.1) converges. But a set of full measure for the product measure does not necessary contain a "rectangle"  $A \times B$  of full measure.

**1.2. Motivation.** The weak disjointness concept appears for the first time in [21] under the name "propriété ergodique produit forte". The aim was to study a well known open problem in pointwise ergodic theory: given two commuting measure preserving transformations  $T$  and  $S$  of the same probability space  $(X, \mathcal{A}, \mu)$ , is it true that for any functions  $f$  and  $g$  in  $L^2(\mu)$ , the sequence

$$\left( \frac{1}{N} \sum_{n=0}^{N-1} f(T^n x) \cdot g(S^n x) \right)_{N>0}$$

converges for  $\mu$ -almost all  $x$  ? (If  $T$  and  $S$  are weakly disjoint, the answer is positive.)

This weak disjointness property is also interesting to study for the following reasons.

- It defines a new invariant in the theory of metric isomorphisms of dynamical systems. (If  $T$  and  $S$  are weakly disjoint and if  $T'$  is a measure theoretic factor of  $T$ , then  $T'$  and  $S$  are weakly disjoint.)
- It has strong links with the rich theory of joinings in Ergodic Theory.
- It gives an opportunity to describe an interesting variety of examples.

**1.3. Brief description of the content.** Let us give a few useful definitions.

**Definitions.** If a dynamical system is weakly disjoint from itself, we say that this dynamical system is *self-weakly disjoint*. A natural generalization of the notion of weak disjointness of a pair of dynamical systems is the notion of weak disjointness of a finite family of dynamical systems (see the discussion at the end of Section 2). If  $k$  copies of a given dynamical systems are weakly disjoint, we say that this dynamical system is *self-weakly disjoint of order  $k$* . Finally a dynamical system is called *universal* if it is weakly disjoint from any dynamical system.

We prove in Section 2 that, if  $T$  and  $S$  are disjoint, then they are weakly disjoint. This already gives a great number of examples. Here are some other ones.

- It is very easy to see that every discrete spectrum dynamical system is universal. This is still true for quasi-discrete spectrum dynamical systems (see 4.1).
- As a direct consequence of del Junco-Keane's study of generic points in the Cartesian square of Chacon's dynamical system ([4]), we observe that the Chacon's dynamical system is self-weakly disjoint. We prove in Section 4.2 that this dynamical system is in fact universal.
- As a direct consequence of Ratner's study of pointwise properties of unipotent transformations (see [20] and the survey [10]), we observe that unipotent transformations are self-weakly disjoint. We show in Section 4.3 that these transformations are universal.
- On the other hand we prove that two dynamical systems with positive entropy are never weakly disjoint, and we give several constructions of zero entropy dynamical systems which are not self-weakly disjoint, including some systems with minimal self-joinings (Section 5).

We prove that Chacon's dynamical system and unipotent transformations are universal as consequences of some general results stated in Section 3, where we describe links between disjointness of dynamical systems and existence of common factors. We use the notion of relative disjointness of two dynamical systems and we obtain two results on weak disjointness.

- If an ergodic dynamical system is weakly disjoint from any ergodic joining of a finite number of copies of itself, then it is weakly disjoint from any ergodic dynamical system.
- If an ergodic dynamical system is self-weakly disjoint of all orders then it is universal.

Thanks to these results, it is also possible to prove that the symbolic dynamical system associated to the Morse sequence is weakly disjoint from any ergodic dynamical system ([17]). Let us also note here that we know an example of an ergodic isometric extension of a discrete spectrum dynamical system, which is not self weakly disjoint ([17]).

**1.4. Questions.** Let  $T$  and  $S$  be two dynamical systems. If  $T$  is ergodic and weakly disjoint from any ergodic component of  $S$ , are  $T$  and  $S$  weakly disjoint? (We know an example of a dynamical system which is weakly disjoint from any of its ergodic components, but which is not self-weakly disjoint, cf. Section 4.4 and [17].)

If a dynamical system is self-weakly disjoint, is it necessarily universal?

We give in this article a positive answer to this question for simple dynamical systems (Section 4.2).

**1.5. Acknowledgments.** We had fruitfull discussions on the subject of weak disjointness with several mathematicians. We thank particularly Ahmed Bouziad, Mariusz Lemańczyk, Christian Mauduit and Anthony Quas who gave us precise contributions which appear in this article.

## 2. MAXIMAL INEQUALITY, DISJOINTNESS AND WEAK DISJOINTNESS

The classical ergodic maximal inequality is used to show that the weak disjointness property can be tested on dense sets of functions.

**Proposition 2.1.** *Let  $(X, \mathcal{A}, \mu, T)$  and  $(Y, \mathcal{B}, \nu, S)$  be two dynamical systems. A sufficient (and obviously necessary) condition for these systems to be weakly disjoint is the following: there exist a dense subset  $F$  of  $L^2(\mu)$  and a dense subset  $G$  of  $L^2(\nu)$ , such that, for any  $f \in F$  and any  $g \in G$ , there exist a set  $A$  in  $\mathcal{A}$  and a set  $B$  in  $\mathcal{B}$  such that*

- $\mu(A) = \nu(B) = 1$
- for all  $x \in A$  and for all  $y \in B$ , the sequence

$$\left( \frac{1}{N} \sum_{n=0}^{N-1} f(T^n x) \cdot g(S^n y) \right)_{N>0}$$

converges.

We say that a dynamical system  $(X, \mathcal{A}, \mu, T)$  is *regular* if  $X$  is a compact metric space, equipped with its Borel  $\sigma$ -algebra  $\mathcal{A}$ , a regular probability measure  $\mu$  and a continuous transformation  $T$ . It is well known that any dynamical system is metrically isomorphic to a regular one (see e.g. [9]).

If  $E$  is a set and  $e$  an element of  $E$ , we denote by  $\delta(e)$  the measure on  $E$  which is the Dirac mass at point  $e$ .

Let  $(X, \mathcal{A}, \mu, T)$  and  $(Y, \mathcal{B}, \nu, S)$  be two regular dynamical systems. The algebras  $\mathcal{C}(X)$  and  $\mathcal{C}(Y)$  of continuous functions on these spaces, equipped with the topology of uniform convergence, are separable. Let  $F$  and  $G$  be countable dense subsets of  $\mathcal{C}(X)$  and  $\mathcal{C}(Y)$  respectively. Using the fact that  $F$  and  $G$  are dense in, respectively,  $L^2(\mu)$  and  $L^2(\nu)$ , and the fact that the set  $\{f \otimes g : f \in F, g \in G\}$  generates a dense linear subspace of  $\mathcal{C}(X \times Y)$  we deduce from Proposition 2.1 the following corollary.

**Corollary 2.2.** *Two regular dynamical systems  $(X, \mathcal{A}, \mu, T)$  and  $(Y, \mathcal{B}, \nu, S)$  are weakly disjoint if and only if there exist  $X_0 \in \mathcal{A}$  and  $Y_0 \in \mathcal{B}$  such that  $\mu(X_0) = \nu(Y_0) = 1$  and, for all  $(x, y) \in X_0 \times Y_0$ , the sequence of probability measures (on  $X \times Y$ )*

$$\left( \frac{1}{N} \sum_{n=0}^{N-1} \delta((T^n x, S^n y)) \right)_{N>0}$$

is weakly convergent.

In the sequel of this article we will use the following notation:

$$\Delta_N(x, y) := \frac{1}{N} \sum_{n=0}^{N-1} \delta((T^n x, S^n y)) .$$

Let us recall some definitions.

A *joining* of two dynamical systems  $(X, \mathcal{A}, \mu, T)$  and  $(Y, \mathcal{B}, \nu, S)$  is a  $T \times S$ -invariant probability measure  $\lambda$  on the product space  $(X \times Y, \mathcal{A} \otimes \mathcal{B})$  whose projections on  $X$  and  $Y$  are  $\mu$  and  $\nu$  respectively. We will also use the word joining to designate the dynamical system  $(X \times Y, \mathcal{A} \otimes \mathcal{B}, \lambda, T \times S)$ . The product measure  $\mu \otimes \nu$  is always a joining.

Two dynamical systems are *disjoint* if the product measure is their only joining. This notion has been introduced and studied by Furstenberg in [8].

Let  $(X, \mathcal{A}, \mu, T)$  be a regular dynamical system. A point  $x$  in  $X$  is called  $(\mu, T)$ -*generic* if the sequence of probability measures  $(\frac{1}{N} \sum_{n=0}^{N-1} \delta(T^n x))$  converges weakly to  $\mu$ . From the Birkhoff ergodic theorem and the separability of the space of continuous functions on  $X$ , we deduce that, if  $T$  is ergodic, then the set  $X_0$  of generic points has full measure. Let  $(X, \mathcal{A}, \mu, T)$  and  $(Y, \mathcal{B}, \nu, S)$  be two ergodic regular dynamical systems and let  $X_0$  and  $Y_0$  be the sets of generic points in each of these systems. If  $x \in X_0$  and  $y \in Y_0$ , then any weak limit point of the sequence of probabilities  $(\Delta_N(x, y))_{N>0}$  is a joining of the two systems. Hence, if there is at most one joining, this sequence converges. (Recall that on a compact metric space, the set of Borel probabilities equipped with the topology of weak convergence is compact metrizable.)

Using the fact that any dynamical system has a regular model we obtain the following consequence of Corollary 2.2.

**Corollary 2.3.** *If two ergodic dynamical systems are disjoint, then they are weakly disjoint.*

*Proof of Proposition 2.1.* We will use the “weak-(1,1) ergodic maximal inequality”: for all  $h \in L^1(\mu)$  and all  $\epsilon > 0$ ,

$$\mu \left( \left\{ x \in X : \sup_{N>0} \frac{1}{N} \left| \sum_{n=0}^{N-1} h(T^n x) \right| > \epsilon \right\} \right) \leq \frac{1}{\epsilon} \|h\|_1 .$$

Let us suppose that the condition stated in Proposition 2.1 is satisfied and consider  $f \in L^2(\mu)$ ,  $g \in L^2(\nu)$ . We fix a sequence  $(f_j)$  in  $F$  which converges to  $f$  in  $L^2(\mu)$  and a sequence  $(g_k)$  in  $G$  which converges to  $g$  in  $L^2(\nu)$ .

The ergodic maximal inequality implies that the sequence

$$\sup_{N>0} \frac{1}{N} \sum_{n=0}^{N-1} |f - f_j|^2 \circ T^n$$

goes to zero in probability when  $j$  goes to infinity. Extracting a subsequence if necessary, we can suppose that this convergence holds almost everywhere. Similarly, we can suppose that

$$\lim_{k \rightarrow \infty} \sup_{N>0} \frac{1}{N} \sum_{n=0}^{N-1} |g - g_k|^2 \circ S^n = 0 \quad \nu - \text{a.e.}$$

There exist subsets of full measure  $A \subset X$  and  $B \subset Y$  such that, for all  $x \in A$  and for all  $y \in B$ ,

$$\lim_{j \rightarrow \infty} \sup_{N>0} \frac{1}{N} \sum_{n=0}^{N-1} |(f - f_j)(T^n x)|^2 = 0 ,$$

$$\lim_{k \rightarrow \infty} \sup_{N>0} \frac{1}{N} \sum_{n=0}^{N-1} |(g - g_k)(S^n y)|^2 = 0 ,$$

$$\text{for all } j, k, \quad \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=0}^{N-1} f_j(T^n x) g_k(S^n y) \quad \text{exists} ,$$

$$\text{for all } j, \quad \sup_{N>0} \frac{1}{N} \sum_{n=0}^{N-1} |f_j(T^n x)|^2 < \infty ,$$

$$\sup_{N>0} \frac{1}{N} \sum_{n=0}^{N-1} |g(S^n y)|^2 < \infty .$$

Using a simple inequality of the type

$$|ab - cd| \leq |(a - a_j)b| + |a_j(b - b_k)| + |a_j b_k - c_j d_k| + |c_j(d_k - d)| + |(c_j - c)d| ,$$

we can write, for any positive integers  $L$ ,  $M$ ,  $j$  and  $k$ ,

$$\begin{aligned} & \left| \frac{1}{L} \sum_{\ell=0}^{L-1} f(T^\ell x)g(S^\ell y) - \frac{1}{M} \sum_{m=0}^{M-1} f(T^m x)g(S^m y) \right| \leq \\ & 2 \sup_{N>0} \left| \frac{1}{N} \sum_{n=0}^{N-1} (f - f_j)(T^n x)g(S^n y) \right| + 2 \sup_{N>0} \left| \frac{1}{N} \sum_{n=0}^{N-1} f_j(T^n x)(g - g_k)(S^n y) \right| + \\ & \left| \frac{1}{L} \sum_{\ell=0}^{L-1} f_j(T^\ell x)g_k(S^\ell y) - \frac{1}{M} \sum_{m=0}^{M-1} f_j(T^m x)g_k(S^m y) \right| . \end{aligned}$$

We choose  $x \in A$  and  $y \in B$ . The last term of the preceding sum goes to zero when  $L$  and  $M$  go to infinity. Using Cauchy-Schwartz inequality, we obtain

$$\begin{aligned} & \limsup_{L, M \rightarrow \infty} \left| \frac{1}{L} \sum_{\ell=0}^{L-1} f(T^\ell x)g(S^\ell y) - \frac{1}{M} \sum_{m=0}^{M-1} f(T^m x)g(S^m y) \right| \leq \\ & 2 \left( \sup_{N>0} \frac{1}{N} \sum_{n=0}^{N-1} |(f - f_j)(T^n x)|^2 \right)^{1/2} \left( \sup_{N>0} \frac{1}{N} \sum_{n=0}^{N-1} |g(S^n y)|^2 \right)^{1/2} + \\ & 2 \left( \sup_{N>0} \frac{1}{N} \sum_{n=0}^{N-1} |f_j(T^n x)|^2 \right)^{1/2} \left( \sup_{N>0} \frac{1}{N} \sum_{n=0}^{N-1} |(g - g_k)(S^n y)|^2 \right)^{1/2} . \end{aligned}$$

Thanks to the condition given on  $A$  and  $B$ , this quantity can be made arbitrarily small for well chosen  $j$  and  $k$ . The Cauchy criteria gives the desired conclusion.  $\square$

The definition of the weak disjointness of a finite family of  $k (\geq 3)$  dynamical systems is a straightforward generalization of the Definition of the Introduction, except that we have to take functions in  $L^k$  of each probability space, in order to have a natural use of Hölder's inequality and the proper extension of Proposition 2.1. Of course, in the case of regular dynamical systems the characterization given by Corollary 2.2 extends straightforwardly to the case of more than two systems.

### 3. COMMON FACTORS AND RELATIVE DISJOINTNESS

**3.1. Some facts about factors and joinings.** A *joining* of a countable family of dynamical systems is a measure on the Cartesian product of these spaces, whose marginals are the given measures and which is invariant under the product transformation.

A *factor* of a dynamical system  $(Y, \mathcal{B}, \nu, S)$  is a sub- $\sigma$ -algebra  $\mathcal{F}$  of  $\mathcal{B}$  which is stable under  $S$ , i.e. which satisfies, for any  $F \in \mathcal{F}$ ,  $S^{-1}F \in \mathcal{F}$ . Let  $(X, \mathcal{A}, \mu, T)$

and  $(Y, \mathcal{B}, \nu, S)$  be two dynamical systems and  $\mathcal{F}$  be a factor of  $S$ . We say that  $\mathcal{F}$  is a *common factor* of  $S$  and  $T$  if there exists a joining  $\lambda$  of  $T$  and  $S$  such that

$$\{\emptyset, X\} \otimes \mathcal{F} \subset \mathcal{A} \otimes \{\emptyset, Y\} \pmod{\lambda}.$$

(If  $\mathcal{C}$  and  $\mathcal{D}$  are two sub- $\sigma$ -algebras of  $\mathcal{A} \otimes \mathcal{B}$ , we write “ $\mathcal{C} \subset \mathcal{D} \pmod{\lambda}$ ” if, for any  $C$  in  $\mathcal{C}$ , there exists  $D$  in  $\mathcal{D}$  such that  $\lambda(C \Delta D) = 0$ . We write “ $\mathcal{C} = \mathcal{D} \pmod{\lambda}$ ” if “ $\mathcal{C} \subset \mathcal{D} \pmod{\lambda}$ ” and “ $\mathcal{D} \subset \mathcal{C} \pmod{\lambda}$ ”.)

Note that if  $\mathcal{F}$  is a common factor of  $S$  and  $T$ , then there exists a sub- $\sigma$ -algebra  $\mathcal{G}$  of  $\mathcal{A}$  such that

$$\{\emptyset, X\} \otimes \mathcal{F} = \mathcal{G} \otimes \{\emptyset, Y\} \pmod{\lambda}.$$

We say that the joining  $\lambda$  *identifies* the  $\sigma$ -algebras  $\mathcal{F}$  and  $\mathcal{G}$ .

Let  $\mathcal{F}$  be a common factor of  $S$  and  $T$ . Let  $\lambda$  be a joining and  $\mathcal{G}$  be a  $\sigma$ -algebra as above. The *relatively independent joining* of  $T$  and  $S$  over  $\mathcal{F}$  is the joining denoted by  $\mu \otimes_{\mathcal{F}} \nu$  and defined by

$$\mu \otimes_{\mathcal{F}} \nu (A \times B) := \int_{X \times Y} \mathbb{P}_{\mu} [A | \mathcal{G}] (x) \cdot \mathbb{P}_{\nu} [B | \mathcal{F}] (y) \, d\lambda(x, y),$$

for  $A \in \mathcal{A}$  and  $B \in \mathcal{B}$ .

Note that, since the two  $\sigma$ -algebras  $\mathcal{F}$  and  $\mathcal{G}$  are identified by the joining  $\lambda$ , we can identify the restriction  $\rho$  of  $\nu$  to  $\mathcal{F}$  with the restriction of  $\mu$  to  $\mathcal{G}$ . Using these identifications we write

$$\mu \otimes_{\mathcal{F}} \nu (A \times B) = \int \mathbb{P}_{\mu} [A | \mathcal{F}] \cdot \mathbb{P}_{\nu} [B | \mathcal{F}] \, d\rho.$$

This formula shows that the relatively independent joining over the common factor does not depend on the choice of the joining  $\lambda$ .

This notion of relatively independent joining of two dynamical systems over a common factor can be extended in a straightforward way to the case of a countable family of dynamical systems.

With this construction of relatively independent joining, it is clear that if two dynamical systems have a common non trivial factor, then they are not disjoint. The reverse is known to be false ([23]) but we have the following result, which can be found in [14].

**Theorem 3.1.** *If the dynamical systems  $T$  and  $S$  are not disjoint, then  $S$  has a non trivial common factor with a joining of a countable family of copies of  $T$ .*

We give a sketch of a proof of this theorem which will be used in the sequel of this article. Let  $\lambda$  be a joining of  $S$  and  $T$ , distinct from the product measure. We consider the relatively independent joining of a countable family of copies of the dynamical system  $(Y \times X, \mathcal{B} \otimes \mathcal{A}, \lambda, S \times T)$  over their common factor  $(Y, \mathcal{B}, \nu, S)$ . This joining is naturally seen as the probability  $\lambda_{\infty}$  on the space  $Y \times X^{\mathbb{N}}$ , which is invariant under the transformation  $S \times T \times T \times T \times \dots$ , and which is defined by

$$\lambda_{\infty}(B \times A_0 \times A_1 \times \dots \times A_k \times X \times X \times \dots) = \int_B \mathbb{P}_{\lambda} [A_0 | \mathcal{B}] \cdot \mathbb{P}_{\lambda} [A_1 | \mathcal{B}] \cdots \mathbb{P}_{\lambda} [A_k | \mathcal{B}] \, d\nu,$$

for  $B \in \mathcal{B}$  and  $A_0, A_1, \dots, A_k \in \mathcal{A}$ . This probability  $\lambda_{\infty}$  is invariant under the shift transformation on each  $y$ -fiber,  $(y, x_0, x_1, x_2, \dots) \mapsto (y, x_1, x_2, x_3, \dots)$ , and on each fiber it is like a product measure. A relative version of Kolmogorov 0-1 law ([14], Lemma 9) gives us that, modulo  $\lambda_{\infty}$ , the  $\sigma$ -algebra of shift-invariant events coincides with  $\mathcal{B} \otimes \{\emptyset, X^{\mathbb{N}}\}$ .



Consider now a bounded measurable function  $f$  on  $X$ , and denote  $f_1(y, x) := f(x)$ ,  $f_\infty(y, (x_k)) := f(x_0)$  for  $y \in Y$ ,  $x \in X$  and  $(x_k) \in X^\mathbb{N}$ . Applying the Birkhoff ergodic theorem in the dynamical system  $(Y \times X^\mathbb{N}, \lambda_\infty, \text{shift})$  we obtain, for  $\lambda_\infty$ -almost all  $(y, (x_k))$ ,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} f(x_k) = \mathbb{E}_{\lambda_\infty} [f_\infty | \mathcal{B} \otimes \{\emptyset, X^\mathbb{N}\}] = \mathbb{E}_\lambda [f_1 | \mathcal{B} \otimes \{\emptyset, X\}] .$$

Since the joining  $\lambda$  is not the product measure, we can choose  $f$  such that the function  $\mathbb{E}_\lambda [f_1 | \mathcal{B} \otimes \{\emptyset, X\}]$  is not constant modulo  $\lambda$ . The factor of  $S$  generated by this function is not trivial, and can be identified, modulo  $\lambda_\infty$ , to a factor of a joining of countably many copies of  $T$ .

**3.2.  $T$ -factors and relative disjointness.** If  $(X, \mathcal{A}, \mu, T)$  and  $(Y, \mathcal{B}, \nu, S)$  are two dynamical systems, we call a  $T$ -factor of  $S$  any common factor of  $S$  with a joining of countably many copies of  $T$ . Theorem 3.1 says that if  $S$  and  $T$  are not disjoint, then  $S$  has a non-trivial  $T$ -factor. In fact, the proof of this theorem gives a more precise result: for any joining  $\lambda$  of  $S$  and  $T$ , for any bounded measurable function  $f$  on  $X$ , the factor of  $S$  generated by the function  $\mathbb{E}_\lambda [f(x) | \mathcal{B}]$  ( $:= \mathbb{E}_\lambda [f_1 | \mathcal{B} \otimes \{\emptyset, X\}]$ ) is a  $T$ -factor of  $S$ . This allows us to give the following extension of the previous theorem.

**Theorem 3.2.** *Given two dynamical systems  $(X, \mathcal{A}, \mu, T)$  and  $(Y, \mathcal{B}, \nu, S)$ , there exists a maximal  $T$ -factor of  $S$ , denoted by  $\mathcal{F}_T$ .*

*Under any joining  $\lambda$  of  $T$  and  $S$ , the  $\sigma$ -algebras  $\mathcal{A} \otimes \{\emptyset, Y\}$  and  $\{\emptyset, X\} \otimes \mathcal{B}$  are conditionally independent given the  $\sigma$ -algebra  $\{\emptyset, X\} \otimes \mathcal{F}_T$ .*

We can say that  $T$  and  $S$  are relatively disjoint over the maximal  $T$ -factor of  $S$ . The proof of the theorem is based on the two following lemmas.

**Lemma 3.3.** *Let  $(B_i)_{i \in I}$  be a countable family of events in  $\mathcal{B}$ , such that for all  $i$  there exists a  $T$ -factor  $\mathcal{F}_i$  of  $S$  containing  $B_i$ . Then there exists a  $T$ -factor of  $S$  containing all the  $B_i$ 's.*

*Proof.* For each  $i \in I$ , we have a joining  $\lambda_i$  of  $S$  with a countable family  $(T_{i,n})_{n \in \mathbb{N}}$  of copies of  $T$ , such that  $\mathcal{F}_i \subset \bigotimes_{n \in \mathbb{N}} \mathcal{A}_{i,n} \text{ mod. } \lambda_i$ . Let us denote by  $Z_i$  the dynamical system defined by  $\lambda_i$ , and by  $\lambda$  the relatively independent joining of all the  $Z_i$ ,  $i \in I$ , over their common factor  $S$ . We can view  $\lambda$  as a joining of  $S$  with the countable family  $(T_{i,n})_{(i,n) \in I \times \mathbb{N}}$  and, for each  $i$  we have

$$\mathcal{F}_i \subset \bigotimes_{(i,n) \in I \times \mathbb{N}} \mathcal{A}_{i,n} \text{ mod. } \lambda.$$

We conclude that the factor of  $S$  generated by all the  $\mathcal{F}_i$ 's is a  $T$ -factor, which certainly contains all the  $B_i$ 's.  $\square$

**Lemma 3.4.** *Let  $\mathcal{F}$  be a factor of  $S$ . If there exists a joining  $\lambda$  of  $T$  and  $S$  under which the  $\sigma$ -algebras  $\mathcal{A} \otimes \{\emptyset, Y\}$  and  $\{\emptyset, X\} \otimes \mathcal{B}$  are not conditionally independent given  $\{\emptyset, X\} \otimes \mathcal{F}$ , then there exists a  $T$ -factor  $\mathcal{F}'$  of  $S$ , not contained in  $\mathcal{F}$ .*

*Proof.* The hypothesis of the lemma implies the existence of a bounded measurable function  $f$  on  $X$  such that, on a set of positive  $\nu$ -measure,

$$\mathbb{E}_\lambda [f(x) | \mathcal{B}] \neq \mathbb{E}_\lambda [f(x) | \mathcal{F}].$$

The factor  $\mathcal{F}'$  of  $S$  generated by the function  $\mathbb{E}_\lambda[f(x)|\mathcal{B}]$  is not contained in  $\mathcal{F}$ . But we saw in the proof of Theorem 3.1 that  $\mathcal{F}'$  is a  $T$ -factor.  $\square$

*Proof of Theorem 3.2.* In order to prove the existence of a maximal  $T$ -factor, we define

$$\mathcal{F}_T := \{B \in \mathcal{B} : B \text{ belongs to a } T\text{-factor of } S\},$$

and we claim that it is a  $T$ -factor. Since  $(Y, \mathcal{B}, \nu)$  is a Lebesgue space, the  $\sigma$ -algebra  $\mathcal{B}$  equipped with the metric  $d(B, C) := \nu(B \Delta C)$  is separable (of course, we identify subsets  $B$  and  $C$  of  $Y$  when  $\nu(B \Delta C) = 0$ ). There exists a countable family  $(B_i)_{i \in I}$  dense in  $\mathcal{F}_T$ , and, thanks to Lemma 3.3, there exists a  $T$ -factor  $\mathcal{F}$  containing all the  $B_i$ 's. By density, we have  $\mathcal{F}_T \subset \mathcal{F}$  but, since  $\mathcal{F}_T$  contains all the  $T$ -factors, we have  $\mathcal{F}_T = \mathcal{F}$ . This proves the first assertion of Theorem 3.2. The second one is just the application of Lemma 3.4 to this factor  $\mathcal{F} = \mathcal{F}_T$ .  $\square$

### 3.3. $T$ -factors and weak disjointness.

**Theorem 3.5.** *If an ergodic dynamical system  $T$  is weakly disjoint from any ergodic joining of a finite family of copies of itself, then it is weakly disjoint from any other ergodic dynamical system  $S$ .*

This theorem can be applied for example to the dynamical system associated to the Morse sequence (see Section 4.4).

*Proof.* Let  $(X, \mathcal{A}, \mu, T)$  be an ergodic dynamical system weakly disjoint from any ergodic joining of a finite family of copies of itself. We observe that this system is also weakly disjoint from any ergodic joining of a countable family of copies of itself, since on  $X^\mathbb{N}$  equipped with any measure  $\lambda$  the set of those  $f$  in  $L^2(\lambda)$  which depend only on finitely many coordinates is dense in  $L^2(\lambda)$ , so we can use Proposition 2.1.

We choose a regular model for the dynamical system  $T$  and we consider another ergodic regular dynamical system  $(Y, \mathcal{B}, \nu, S)$ . We denote by  $\mathcal{F}_T$  the maximal  $T$ -factor of  $S$ . We fix a countable dense set  $D$  of continuous functions on  $Y$ , and for each  $g \in D$ , we fix a version of the conditional expectation  $\mathbb{E}[g|\mathcal{F}_T]$ .  $\mathcal{F}_T$  is also a factor of a countable joining  $\tau$  of  $T$ . Since  $\mathcal{F}_T$  is ergodic (because it is a factor of  $S$ ), it is also a factor of almost every ergodic component of  $\tau$ , and since  $T$  is ergodic, almost every ergodic component of  $\tau$  is a joining of  $T$ . Hence  $\mathcal{F}_T$  is a factor of an ergodic countable joining of  $T$ . Weak-disjointness passes to factors, so  $\mathcal{F}_T$  is weakly disjoint from  $T$ . Then there exist sets of full measure  $X_0$  and  $Y_0$  of generic points in each of the dynamical systems such that, for any  $f \in \mathcal{C}(X)$  and any  $g \in D$ , for all  $x \in X_0$  and all  $y \in Y_0$ ,

$$(3.1) \quad \text{the sequence} \quad \left( \frac{1}{N} \sum_{n < N} f(T^n x) \cdot \mathbb{E}[g|\mathcal{F}_T](S^n y) \right) \quad \text{converges}.$$

Let  $x \in X_0$  and  $y \in Y_0$ . Let  $\lambda$  and  $\lambda'$  be two weak limit values of the sequence of probabilities

$$(3.2) \quad \Delta_N(x, y) = \frac{1}{N} \sum_{n < N} \delta((T^n x, S^n y)).$$

From (3.1) we deduce that the measures  $\lambda$  and  $\lambda'$  coincide on the  $\sigma$ -algebra  $\mathcal{A} \otimes \mathcal{F}_T$ . By the genericity condition both of these measures are joinings of  $T$  and  $S$ , hence, by Theorem 3.2, the  $\sigma$ -algebras  $\mathcal{A} \otimes \{\emptyset, Y\}$  and  $\{\emptyset, X\} \otimes \mathcal{B}$  are conditionally independent

given the  $\sigma$ -algebra  $\{\emptyset, X\} \otimes \mathcal{F}_T$ . We claim that this implies that  $\lambda = \lambda'$ , and this is sufficient to establish the convergence of (3.2) and the weak disjointness of  $T$  and  $S$ . The claim can be justified by the following basic lemma, applied to  $\Omega = X \times Y$ ,  $\alpha = \mathcal{A} \otimes \{\emptyset, Y\}$ ,  $\beta = \{\emptyset, X\} \otimes \mathcal{B}$  and  $\gamma = \{\emptyset, X\} \otimes \mathcal{F}_T$ . The proof of this lemma is left to the reader.

**Lemma 3.6.** *Let  $\Omega$  be a set and  $\alpha, \beta, \gamma$  be three  $\sigma$ -algebras of subsets of  $\Omega$ . Let  $\lambda$  and  $\lambda'$  be two probability measures on  $(\Omega, \alpha \vee \beta \vee \gamma)$ . If  $\lambda$  and  $\lambda'$  coincide on  $\alpha \vee \gamma$  and coincide on  $\beta \vee \gamma$ , and if under each of these two measures,  $\alpha$  and  $\beta$  are conditionally independent given  $\gamma$ , then  $\lambda$  and  $\lambda'$  coincide on  $\alpha \vee \beta$ .*

□

We don't know if, under the hypothesis of Theorem 3.5, it is possible to conclude that  $T$  is universal (that is to say if it is possible to remove the ergodicity condition on  $S$ ). In order to get around this difficulty, we introduce a stronger hypothesis on  $T$ .

It is clear from our preceding discussions that a regular system  $(X, \mathcal{A}, \mu, T)$  is self-weakly disjoint of order  $k \geq 2$  if and only if there exists  $X_0 \in \mathcal{A}$ , of full measure, such that, for any  $(x_1, x_2, \dots, x_k) \in X_0^k$  the sequence of probabilities

$$\Delta_N(x_1, x_2, \dots, x_k) := \frac{1}{N} \sum_{n=0}^{N-1} \delta((T^n x_1, T^n x_2, \dots, T^n x_k))$$

converges weakly on the space  $X^k$ .

**Theorem 3.7.** *If an ergodic dynamical system is self-weakly disjoint of all orders, then it is universal.*

This theorem will be applied to Chacon's dynamical system (Section 4.2) and to unipotent transformations (Section 4.3).

We will use the following lemma.

**Lemma 3.8.** *If  $T$  and  $S$  are two continuous transformations of the compact metric spaces  $X$  and  $Y$ , and if  $X_0$  is a Borel subset of  $X$ , then the set*

$$C_{X_0} := \{y \in Y : \forall x \in X_0, (\Delta_N(x, y)) \text{ converges}\}$$

*is universally measurable in  $Y$ .*

*Proof of Lemma 3.8.* The complement of  $C_{X_0}$  in  $Y$  is the projection of the set

$$\{(x, y) \in X_0 \times Y : (\Delta_N(x, y)) \text{ does not converge}\},$$

onto  $Y$ . This set is a Borel subset of  $X \times Y$ , and its projection onto  $Y$  is analytic in  $Y$ , hence universally measurable (see e.g. [12]). □

*Proof of Theorem 3.7.* We consider an ergodic regular system  $(X, \mathcal{A}, \mu, T)$ , self-weakly disjoint of all orders. There exists a Borel subset  $X_0$  of  $X$ , with  $\mu(X_0) = 1$ , such that, for  $k \geq 2$  and  $x_1, x_2, \dots, x_k \in X_0$ , the sequence  $(\Delta_N(x_1, x_2, \dots, x_k))$  converges. We can suppose also that all points in  $X_0$  are generic in the dynamical system. For any joining  $\lambda$  of  $k$  copies of  $T$ , we have  $\lambda(X_0^k) = 1$ . This implies that  $T$  is weakly disjoint from any joining of countably many copies of  $T$ . We have even more: for any joining  $(Z, \mathcal{C}, \lambda, U)$  of countably many copies of  $T$ , there exist  $Z_0$  subset of  $Z$  of full measure such that the sequence  $(\Delta_N(x, z))$  converges for all

$x \in X_0$  and all  $z \in Z_0$ . Following the proof of Theorem 3.5, this means that our set  $X_0$  will work for any choice of the ergodic dynamical system  $S$ .

Let now  $(Y, \mathcal{B}, \nu, S)$  be any dynamical system, not necessarily ergodic. We write  $\nu = \int_{\mathcal{E}_S} \eta dP(\eta)$  the ergodic disintegration of  $\nu$ , where  $P$  is a probability measure on the set  $\mathcal{E}_S$  of  $S$ -invariant ergodic probability measures on  $Y$ . For  $\eta \in \mathcal{E}_S$ , there exists  $Y_\eta$  of full  $\nu$ -measure in  $Y$  such that, for all  $x \in X_0$  and all  $y \in Y_\eta$ , the sequence  $(\Delta_N(x, y))$  converges. But Lemma 3.8 tells us that the set

$$Y_0 := \{y \in Y : \forall x \in X_0, (\Delta_N(x, y)) \text{ converges}\}$$

is  $\nu$ -measurable. Since we have  $\eta(Y_0) = 1$  for all  $\eta \in \mathcal{E}_S$ , we conclude that  $\nu(Y_0) = 1$ , which implies that  $T$  and  $S$  are weakly disjoint.  $\square$

#### 4. EXAMPLES OF WEAK DISJOINTNESS

**4.1. Quasi-discrete spectrum dynamical systems.** Let  $(X, \mathcal{A}, \mu, T)$  be a discrete spectrum dynamical system. The family of  $T$ -eigenfunctions in  $L^2(\mu)$  generates a dense linear subspace of  $L^2(\mu)$ . Let  $f$  be a  $T$ -eigenfunction; there exists  $\lambda \in \mathbb{C}$  such that  $f \circ T = \lambda f$ . Let  $(Y, \mathcal{B}, \nu, S)$  be another dynamical system and  $g \in L^2(\nu)$ . The Birkhoff ergodic theorem applied to the product of the systems gives us that, for  $\mu \otimes \nu$ -almost all  $(x, y)$ , the sequence

$$f(x) \frac{1}{N} \sum_{n=0}^{N-1} \lambda^n g(S^n y) = \frac{1}{N} \sum_{n=0}^{N-1} f(T^n x) g(S^n y)$$

converges. But the convergence does not depend on  $x$ . Thus we have : for  $\nu$ -almost all  $y$ , for all  $x$ , the sequence (1.1) converges. By density, and by Proposition 2.1, this implies that  $T$  and  $S$  are weakly disjoint. We have proved that any discrete spectrum dynamical system is universal.

This argument can be extended to quasi-discrete spectrum dynamical systems, that is to say, dynamical systems in which the generalized eigenfunctions, in the sense of [11], generates a dense linear subspace of  $L^2$ .

Let us recall this definition more precisely. Let  $(X, \mathcal{A}, \mu, T)$  be a dynamical system. We denote by  $E_0$  the set of constant complex functions and we define by induction an increasing sequence  $(E_k)_{k \geq 0}$ , of subsets of  $L^2(\mu)$  by

$$E_{k+1} := \{f \in L^2(\mu) : f \circ T = g \cdot f, \text{ with } g \in E_k, |g| = 1\}.$$

We say that  $T$  has quasi-discrete spectrum if  $\cup_{k \geq 0} E_k$  generates a dense linear subspace of  $L^2(\mu)$ .

**Proposition 4.1.** *Any quasi-discrete spectrum dynamical system is universal.*

*Proof.* By Proposition 2.1, it is enough to study the convergence of (1.1) when  $f$  belongs to one of the  $E_k$ 's. If  $f \in E_k$ , then  $f(T^n x) = f(x) \exp(iR_x(n))$  where  $R_x$  is a real polynomial depending on the point  $x$ . The conclusion follows from the following result, which can be found in [15].

**Proposition.** *Let  $(Y, \mathcal{B}, \nu, S)$  be a dynamical system. For all  $g \in L^1(\nu)$ , for  $\nu$ -almost all  $y$ , for all real polynomial  $P$  and all continuous periodic real function  $\phi$  on  $\mathbb{R}$ , the sequence  $(\frac{1}{N} \sum_{n < N} \phi(P(n)) \cdot g(S^n y))$  converges.*

$\square$

**4.2. Chacon's transformation.** The Chacon dynamical system is a well known example of a weakly mixing but not mixing system. We denote by  $(X, \mathcal{A}, \mu, T)$  an invertible regular version of this system. From the result proved in [4] we deduce directly the existence of a subset  $X_0$  of full measure in  $X$  such that every point in  $X_0$  is generic, and such that for all  $x$  and  $y$  in  $X_0$ , either there exist  $p \in \mathbb{Z}$  such that  $T^p x = y$ , or  $\Delta_N(x, y) \longrightarrow \mu \otimes \mu$ , when  $N \rightarrow \infty$ . Of course, if  $x$  is generic and if  $T^p x = y$ , then  $\Delta_N(x, y)$  goes to the image under  $Id \times T^p$  of the diagonal measure on  $X^2$ .

Therefore this dynamical system is self-weakly disjoint. In fact, we have more.

**Proposition 4.2.** *The Chacon dynamical system is universal.*

*Proof.* By Theorem 3.7, it is sufficient to prove that the Chacon dynamical system is self-weakly disjoint of all orders.

Let  $k$  be an integer  $\geq 2$ , and  $x_1, x_2, \dots, x_k$  in  $X_0$ . Let  $\lambda$  and  $\lambda'$  be two limit values of the sequence  $(\Delta_N(x_1, x_2, \dots, x_k))$ .

Since every point in  $X_0$  is generic,  $\lambda$  and  $\lambda'$  are two joinings of  $k$  copies of  $T$ . Since for all  $(x, y) \in (X_0)^2$  the sequence  $(\Delta_N(x, y))$  converges, the restrictions of  $\lambda$  and  $\lambda'$  to any sub- $\sigma$ -algebra generated by two coordinates always coincide. (These restrictions are either the product measure  $\mu \otimes \mu$ , or of the form  $\Lambda_p$ , where  $\Lambda_p(A \times B) := \mu(A \cap T^{-p}B)$ .) But we know that Chacon's dynamical system has minimal self-joinings of all orders ([5]), and we conclude that  $\lambda = \lambda'$ . This proves that the system is self-weakly disjoint of all orders.  $\square$

In fact, the proof of Theorem 3.7 shows that, in order to prove that an ergodic dynamical system  $T$  is universal, it is enough to prove the existence of  $X_0 \subset X$ , of full  $\mu$ -measure, such that, for any integer  $k \geq 1$ , and any ergodic joining  $\lambda$  of  $k$  copies of  $T$ , there exists  $Y_0 \subset X^k$ , of full  $\lambda$ -measure, such that, for all  $x \in X_0$  and all  $y \in Y_0$ , the sequence  $(\Delta_N(x, y))$  converges. This remark allows us to extend the preceding argument to all simple dynamical systems.

Let us recall that an ergodic dynamical system  $(X, \mathcal{A}, \mu, T)$  is *simple* (see [6] or [24]) if, for any positive integer  $k$ , and any ergodic joining  $\lambda$  of  $k$  copies of  $T$ , say  $T_1, T_2, \dots, T_k$ , the index set  $\{1, 2, \dots, k\}$  can be divided into subsets  $E_1, E_2, \dots, E_r$  with the following properties:

- if  $i$  and  $j$  are two indices in the same subset  $E_l$ , then the  $\sigma$ -algebras  $\mathcal{A}_i$  and  $\mathcal{A}_j$  are identified by  $\lambda$ ,
- if we choose an element  $i_l$  of each subset  $E_l$ , the  $\sigma$ -algebras  $\mathcal{A}_{i_1}, \dots, \mathcal{A}_{i_r}$  are independent under  $\lambda$ .

This implies that the dynamical system  $(X^k, \mathcal{A}^{\otimes k}, \lambda, T^{\times k})$  is isomorphic to the  $r$ -th Cartesian power of  $(X, \mathcal{A}, \mu, T)$ .

**Proposition 4.3.** *If a dynamical system is simple and weakly disjoint from itself, then it is universal.*

*Proof.* Let  $(X, \mathcal{A}, \mu, T)$  be a simple dynamical system, weakly disjoint from itself. Thanks to this last property and the ergodicity of  $T$ , there exists  $X_0 \subset X$ , of full measure such that any point in  $X_0$  is generic and such that, for all  $(x, y) \in (X_0)^2$ , the sequence  $(\Delta_N(x, y))$  converges. Let  $\gamma$  be an ergodic joining of finitely many copies of  $T$ . Because of the simplicity of  $T$ , we can suppose that  $\gamma$  is the product measure  $\mu^{\otimes r}$  on  $X^r$ . By the ergodicity of  $\gamma$ , there exists  $Y_0 \subset (X_0)^r$ , with  $\gamma(Y_0) = 1$ , such that, for all  $(x_1, x_2, \dots, x_r) \in Y_0$ ,  $\Delta_N(x_1, x_2, \dots, x_r) \longrightarrow \mu^{\otimes r}$ .

Let  $x \in X_0$  and  $(x_1, x_2, \dots, x_r) \in Y_0$ . We want to prove that the sequence  $(\Delta_N(x, x_1, x_2, \dots, x_r))$  has at most one limit point. This will prove that  $T$  is weakly disjoint from any ergodic joining of a finite number of copies of itself and the conclusion will follow from Theorem 3.7 and the remark above.

Let  $\lambda$  and  $\lambda'$  be two limit points of the sequence  $(\Delta_N(x, x_1, x_2, \dots, x_r))$ . For each  $i$  between 1 and  $r$ , the points  $x$  and  $x_i$  are in  $X_0$ , hence

$$(4.1) \quad \lambda|_{\mathcal{A} \otimes \mathcal{A}_i} = \lambda'|_{\mathcal{A} \otimes \mathcal{A}_i}.$$

From the choice of  $Y_0$ , we deduce that

$$(4.2) \quad \lambda|_{\mathcal{A}_1 \otimes \dots \otimes \mathcal{A}_r} = \lambda'|_{\mathcal{A}_1 \otimes \dots \otimes \mathcal{A}_r} = \mu^{\otimes r}.$$

All the points  $x$  and  $x_i$  being generic, the measures  $\lambda$  and  $\lambda'$  are joinings of  $r+1$  copies of  $T$ . Since  $T$  is simple we can write their ergodic disintegration

$$(4.3) \quad \lambda = \alpha \mu \otimes \mu^{\otimes r} + (1 - \alpha) \int_J \eta \, dP(\eta)$$

$$(4.4) \quad \lambda' = \alpha' \mu \otimes \mu^{\otimes r} + (1 - \alpha') \int_J \eta \, dP'(\eta),$$

where  $\alpha$  and  $\alpha'$  are real numbers between 0 and 1, and  $P, P'$  are two probabilities on the space  $J$  of ergodic  $(r+1)$ -joinings  $\eta$  of  $T$  such that

- $\eta|_{\mathcal{A}_1 \otimes \dots \otimes \mathcal{A}_r} = \mu^{\otimes r}$ ,
- there exists a unique  $i \in \{1, 2, \dots, r\}$  such that  $\mathcal{A} = \mathcal{A}_i \text{ mod } \eta$ .

For  $i \in \{1, 2, \dots, r\}$ , let us denote  $J_i := \{\eta \in J : \mathcal{A} = \mathcal{A}_i \text{ mod } \eta\}$ . the sets  $J_i$  form a partition of  $J$ . A restriction of (4.3) to the  $\sigma$ -algebra  $\mathcal{A} \otimes \mathcal{A}_i$  gives

$$(4.5) \quad \lambda|_{\mathcal{A} \otimes \mathcal{A}_i} = \left( \alpha + (1 - \alpha)P(J \setminus J_i) \right) \mu \otimes \mu + (1 - \alpha) \int_{J_i} \eta|_{\mathcal{A} \otimes \mathcal{A}_i} \, dP(\eta).$$

From (4.1), (4.5) and unicity of the ergodic disintegration, we deduce that

$$(4.6) \quad (1 - \alpha) \int_{J_i} \eta|_{\mathcal{A} \otimes \mathcal{A}_i} \, dP(\eta) = (1 - \alpha') \int_{J_i} \eta|_{\mathcal{A} \otimes \mathcal{A}_i} \, dP'(\eta).$$

But a probability  $\eta \in J_i$  is uniquely determined by its restriction to  $\mathcal{A} \otimes \mathcal{A}_i$ , so (4.6) implies that

$$(1 - \alpha) \int_{J_i} \eta \, dP(\eta) = (1 - \alpha') \int_{J_i} \eta \, dP'(\eta).$$

If there exists  $i$  such that  $P(J_i) > 0$ , then we obtain successively  $\alpha = \alpha'$ ,  $P = P'$  and  $\lambda = \lambda'$ . If for all  $i$ ,  $P(J_i) = 0$ , then  $\lambda = \mu^{\otimes(r+1)}$ , and

$$\lambda'|_{\mathcal{A} \otimes \mathcal{A}_i} = \lambda|_{\mathcal{A} \otimes \mathcal{A}_i} = \mu \otimes \mu;$$

since  $T$  is simple, this gives  $\lambda' = \mu^{\otimes(r+1)}$ . In all the cases we conclude that the limit value of the sequence  $(\Delta_N(x, x_1, x_2, \dots, x_r))$  is unique.  $\square$

**4.3. Unipotent transformations.** We refer to the survey [10] for an introduction to this subject.

Let  $G$  be a connected Lie group and  $A$  its Lie algebra. Let  $\Gamma$  be a lattice in  $G$ , i.e. a discrete subgroup of  $G$  such that the homogeneous space  $G/\Gamma$  has finite Haar volume. To any element  $g$  of  $G$  is associated its adjoint  $\text{Ad}(g)$ , which is a linear operator of  $A$ . The element  $g$  is called *unipotent* if  $\text{Ad}(g)$  has only 1 as eigenvalue. To any  $g$  in  $G$  is associated a dynamical system, which is the translation (on the left) by  $g$  on the homogeneous space  $G/\Gamma$ . This dynamical system is denoted by  $(G, \Gamma, g)$ . This system is called a *unipotent transformation* if  $g$  is unipotent.

One well known example is the horocycle transformation on a riemannian surface of curvature  $-1$ , which can be represented by the translation by  $g = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$  on the quotient of the group  $G = \text{Sl}(2, \mathbb{R})$  by a lattice  $\Gamma$ .

The following theorem is due to M. Ratner ([20]).

**Theorem.** *Let  $(G, \Gamma, g)$  be a unipotent transformation. In this dynamical system, every element is generic for a measure (which may depend on the element). Equivalently, for every  $x \in G/\Gamma$ , the sequence of probability measures  $\left(\frac{1}{N} \sum_{n=0}^{N-1} \delta(g^n x)\right)$  is weakly convergent.*

The class of unipotent transformations is stable under Cartesian products. Thus, by a direct application of Ratner's Theorem, we observe that any unipotent transformation is self-weakly disjoint of all orders. And using Theorem 3.7, we obtain the following result.

**Proposition 4.4.** *Any ergodic unipotent transformation is universal.*

**4.4. The dynamical system associated to the Morse sequence.** In this section we present another example. We announce results that will be described and proved in details in [17].

The Morse (or Prouhet-Thue-Morse) sequence  $u = (u_n)_{n \geq 0}$  is the sequence of 0's and 1's inductively defined by  $u_0 = 0$ ,  $u_{2n} = u_n$  and  $u_{2n+1} = 1 - u_n$ . It admits many other simple descriptions and serves as a typical example for various objects, in combinatorics, number theory, symbolic dynamics and geometry. For some historical comments and a large list of references, we refer to [19].

We consider the space  $\{0, 1\}^{\mathbb{N}}$  of 0-1 sequences equipped with the product topology and the shift transformation  $\theta$ . The closure of the orbit of  $u$  under  $\theta$  in this compact space is denoted by  $K$ :

$$K := \overline{\{(u_{n+k})_{n \geq 0} : k \geq 0\}}.$$

$K$  is the compact set of all sequences of 0's and 1's whose words are words from the Morse sequence. It is known that there exists on  $K$ , equipped with its Borel  $\sigma$ -algebra, a unique  $\theta$ -invariant probability measure, that we denote by  $\mu$ . The dynamical system  $(K, \mu, \theta)$  is ergodic and every point is generic. We call this system the Morse dynamical system. Many things are known on the ergodic and spectral properties of the Morse dynamical system. Note that it can be described as a two point extension of the dyadic odometer (see e.g. [16]).

Relating to the weak disjointness property, we have the following results ([17]).

- (1) For every continuous function  $f$  on the cube  $K^3$  the sequence

$$\left( \frac{1}{N} \sum_{n=0}^{N-1} f \circ (\theta \times \theta \times \theta)^n \right)$$

is everywhere convergent. Consequently, the Morse dynamical system is self-weakly disjoint of order 3 (and hence of order 2).

- (2) If  $A$  is any measurable subset of  $\mu \otimes \mu$  positive measure in  $K^2$ , there exist two elements  $((a_n), (b_n))$  and  $((c_n), (d_n))$  in  $A$  such that the sequence

$$\left( \frac{1}{N} \sum_{n=0}^{N-1} (-1)^{a_n+b_n+c_n+d_n} \right)$$

does not converge. Consequently the Cartesian square of the Morse dynamical system is not self-weakly disjoint. In particular, the Morse dynamical system is not self-weakly disjoint of order 4 (and hence of any order  $\geq 4$ ).

- (3) Denote by  $M^2$  the Cartesian square of the Morse dynamical system. Almost every ergodic component of  $M^2$  is weakly disjoint from  $M^2$ . Consequently, almost every pair of ergodic components of  $M^2$  is a pair of weakly disjoint dynamical systems. But some of these ergodic components are not self weakly disjoint.
- (4) The Morse dynamical system is weakly disjoint from any ergodic joining of finitely many copies of itself. By Theorem 3.5, this implies that the Morse dynamical system is weakly disjoint from any other ergodic dynamical system.

## 5. EXAMPLES OF LACK OF WEAK DISJOINTNESS

The following proposition gives a way of showing that two dynamical systems are not weakly disjoint.

**Proposition 5.1.** *Let  $(X, \mathcal{A}, \mu, T)$  and  $(Y, \mathcal{B}, \nu, S)$  be two dynamical systems. If there exist  $f \in L^2(\mu)$ ,  $g \in L^2(\nu)$  and a measurable map  $\varphi$  from  $X$  into  $Y$  such that  $\varphi^* \mu \ll \nu$  and*

$$\mu \left\{ x \in X : \text{the sequence } \left( \frac{1}{N} \sum_{n=0}^{N-1} f(T^n x) \cdot g(S^n \varphi(x)) \right) \text{ does not converge} \right\} > 0,$$

*then  $T$  and  $S$  are not weakly disjoint.*

*Proof.* The proof is straightforward: if  $A \in \mathcal{A}$  and  $B \in \mathcal{B}$  are such that the averages (1.1) converge for all  $x \in A$  and all  $y \in B$ , then  $\mu(A \cap \varphi^{-1}(B)) < 1$ ; if furthermore  $\mu(A) = 1$ , this implies that  $\mu(\varphi^{-1}(B)) < 1$ , hence  $\nu(B) < 1$ .  $\square$

### 5.1. Positive entropy.

**Proposition 5.2.** *Two dynamical systems of positive entropy are never weakly disjoint.*

Let  $p \in (0, 1/2]$ . We consider the space  $Z = \{-1, 0, 1\}^{\mathbb{N}}$  of sequences on the three letters  $-1$ ,  $0$  and  $1$ . We equip this space with the product measure  $\pi = (p, 1-2p, p)^{\otimes \mathbb{N}}$  and with the shift transformation  $\theta$ . The dynamical system  $(Z, \pi, \theta)$  is a Bernoulli shift of entropy  $2p \ln p + (1-2p) \ln(1-2p)$  which can be fixed arbitrarily small by choosing  $p$  small enough.



**Lemma 5.3.** *The dynamical system  $(Z, \pi, \theta)$  is not self-weakly disjoint.*

*Proof of the Lemma.* We define a transformation  $\varphi$  of  $Z$  by the following rule: if  $z = (z_n)_{n \geq 0} \in Z$  then  $\varphi(z) = (z'_n)_{n \geq 0}$  is given by

$$z'_n = \begin{cases} z_n & \text{if } 2^\ell \leq n < 2^{\ell+1} \text{ with } \ell \text{ even,} \\ -z_n & \text{if } 2^\ell \leq n < 2^{\ell+1} \text{ with } \ell \text{ odd.} \end{cases}$$

The map  $\varphi$  preserves the product measure  $\pi$ . The sequence  $(z_n z'_n)$  takes alternatively the values 1 or 0 and the values  $-1$  or 0 in the successive dyadic blocks of indices. Moreover the asymptotic frequency of non zero terms in this sequence is given by the large law of large numbers : for  $\pi$ -almost all  $z$  this frequency is equal to  $2p > 0$ . Then it is easy to see that the sequence

$$\left( \frac{1}{N} \sum_{n=0}^{N-1} z_n z'_n \right)$$

diverges almost everywhere. We can apply Proposition 5.1 to the function  $f(z) = g(z) = z_0$ , and we have proved Lemma 5.3.  $\square$

*Proof of Proposition 5.2.* By a classical theorem of Sinai, we know that if  $T$  is a dynamical system of entropy  $h > 0$ , then any Bernoulli shift of entropy  $\leq h$  is a factor of  $T$ . If  $T$  and  $S$  have positive entropies, and if  $p$  is small enough, then the dynamical system  $(Z, \pi, \theta)$  is a common factor of  $T$  and  $S$ . Because the weak-disjointness goes to factors, Lemma 5.3 gives directly the Proposition.  $\square$

We cited already an example for a zero entropy dynamical system which is not self-weakly disjoint : the Cartesian square of Morse system. This system is not ergodic. Let us present now three types of constructions of ergodic examples. The two first mimic the Bernoulli case. The third one gives a great variety of rank one transformations.

**5.2. A cutting and stacking procedure.** Here is an abstract of what we want to describe : we consider a cutting and stacking construction of the Bernoulli shift. After each step of this construction we add a new step, just by cutting each tower into two equal pieces and stacking these two pieces. This destroys the entropy but stays close enough to the Bernoulli case. This construction is inspired by a process described in [22]. Let us go into some details.

We consider the Bernoulli scheme on two letters, with uniform probability. Let us recall the cutting and stacking construction of this dynamical system. At the first stage, we have 2 intervals (towers of height 1) of the same size, labelled by 0 and 1 respectively. Then we cut each of these intervals into 4 parts of equal sizes, that we pairwise stack in order to obtain 4 towers of height 2, associated to the labels (0, 0), (0, 1), (1, 0) and (1, 1). This gives the second stage. At stage  $n$ , we have  $2^{2^{n-1}}$  towers of height  $2^{n-1}$ , all with the same size. We cut each of these towers into  $2^{2^{n-1}+1}$  parts. These new towers are pairwise stacked, to obtain  $2^{2^n}$  towers of height  $2^n$ , which are labelled by all the elements of  $\{0, 1\}^{2^n}$ . Our space is the union of the intervals at the beginning. Each point of the space is uniquely determined by the bilateral sequence of 0's and 1's that can be read above and below it in the tower where it appears. The transformation consists only in climbing of one level in the tower. Via this 0 – 1 coding this dynamical system is exactly the  $(\frac{1}{2}, \frac{1}{2})$  Bernoulli scheme.

Now we follow the preceding procedure, but we insert between each stage a simple cutting and stacking of each individual tower of our scheme. (A simple cutting and stacking of a tower consists in stacking two halves of the tower.) So at the first stage, we have 2 towers of height 2, with labels  $(0,0)$  and  $(1,1)$ . At the second stage we have 4 towers of height 8, with labels  $0^8, (0^2 1^2)^2, (1^2 0^2)^2$  and  $1^8$ . At stage  $n$ , we have  $2^{2^{n-1}}$  towers of height  $2^{2n-1}$ .

This procedure defines a measure preserving dynamical system which can be described as a shift invariant probability measure  $\mu$  on the space  $\Omega := \{0,1\}^{\mathbb{N}}$ .

This system has zero entropy. Indeed, if we denote by  $a_n$  the number of words of length  $4^n$  which appear with positive  $\mu$ -measure, it is not difficult to verify that  $a_{n+1} \leq 2^{2n+1} a_n^3$ ; hence we have  $\lim 4^{-n} \ln a_n = 0$ .

Let us prove that this dynamical system is not self-weakly disjoint. Because of the repetition of simple cutting and stacking of towers of height  $2^{2k}$ , almost all sequence of our system can be described as an initial word of length less than  $2^{2k+1}$  followed by a sequence of words of length  $2^{2k+1}$  each of them being a concatenation of two identical words of length  $2^{2k}$ ; for  $\mu$ -almost all  $\omega = (\omega(i))_{i \geq 0} \in \Omega$ , for all  $k \geq 0$ , there exists an integer  $j = j(k, \omega)$  between 0 and  $2^{2k+1} - 1$  which marks the initial place of the repeated blocks of length  $2^{2k}$ . More precisely, for all  $n \geq 0$ , for all  $i$  with  $0 \leq i < 2^{2k}$ , we have

$$\omega(i + j + n2^{2k+1}) = \omega(i + 2^{2k} + j + n2^{2k+1}).$$

The repetition of independent choices of (more and more longer) words insures that, almost surely, for each  $k$ , the integer  $j$  is unique. We call the finite sequence  $(\omega(i + j + n2^{2k+1}))_{0 \leq i < 2^{2k+1}}$  the  $(n, k)$ -word of the sequence  $\omega$ . This word is the concatenation of two identical words of length  $2^{2k}$ .

Given  $n, k > 0$ , the transformation of  $\Omega$  which consists in changing all the letters of the  $(n, k)$ -word and only these ones is an (almost-everywhere defined) involution which preserves the probability measure  $\mu$ . More generally, if  $((n_\ell, k_\ell))_{\ell \geq 0}$  is a sequence of pairs of positive integers such that  $n_\ell 2^{2k_\ell+1} \rightarrow \infty$  when  $\ell \rightarrow \infty$ , then the transformation of  $\Omega$  which consists in changing successively the letters of all the  $(n_\ell, k_\ell)$  words is almost everywhere well defined and measure preserving.

Let us consider an increasing sequence  $(k_\ell)$  of positive integers which goes to  $\infty$  quickly enough (to be made precise later). We denote by  $\varphi$  the transformation of  $\Omega$  which consists in changing the letters of the  $(1, k_\ell)$ -words. Note that the  $(1, k)$ -word always begins after the index  $2^{2k+1}$  and that it ends before the index  $3 \times 2^{2k+1}$ . Hence, if  $3 \times 2^{2k_{\ell-1}+1} \leq i < 2^{2k_\ell+1}$ , then  $\omega_i = (\varphi(\omega))_i$ . Consequently, if  $k_\ell$  is large enough with respect to  $k_{\ell-1}$ , then

$$2^{-(2k_\ell+1)} \sum_{i < 2^{2k_\ell+1}} (-1)^{\omega_i + (\varphi(\omega))_i} > \frac{1}{2}.$$

On the other hand, since  $\omega_i = 1 - (\varphi(\omega))_i$  for the indices  $i$  of the  $(1, k_\ell)$ -word, we have

$$3^{-1} \times 2^{-(2k_\ell+1)} \sum_{i < 3 \times 2^{2k_\ell+1}} (-1)^{\omega_i + (\varphi(\omega))_i} \leq \frac{1}{3};$$

The sequence

$$\left( \frac{1}{n} \sum_{i < n} (-1)^{\omega_i + (\varphi(\omega))_i} \right)$$

does not converge, and the proof is finished by Proposition 5.1.

**5.3. A skew-product construction.** Let  $(X, \mathcal{A}, \mu, T)$  be an ergodic dynamical system, and  $a : X \rightarrow \mathbb{Z}$  a measurable map such that

- (1) there does not exist  $b : X \rightarrow \mathbb{Z}$ , measurable, such that, for  $\mu$ -almost every  $x$ ,  $a(x) = b(Tx) - b(x)$  ;
- (2) for  $\mu$ -almost every  $x$ , there exists  $n > 0$  such that  $\sum_{k=0}^{n-1} a(T^k x) = 0$ .

Let us denote  $(\Omega, \nu, \theta)$  the two-sided Bernoulli scheme :  $\Omega = \{0, 1\}^{\mathbb{Z}}$ ,  $\nu$  is the uniform product measure  $(1/2, 1/2)^{\otimes \mathbb{Z}}$  and  $\theta$  is the shift. We consider the transformation  $T_a$  of the space  $X \times \Omega$  defined by

$$T_a(x, \omega) := (Tx, \theta^{a(x)} \omega) .$$

This transformation preserves the product measure  $\mu \otimes \nu$ .

**Proposition 5.4.** *The dynamical system  $(X \times \Omega, \mu \otimes \nu, T_a)$  is not self-weakly disjoint. If  $(X, \mu, T)$  has zero entropy, then  $(X \times \Omega, \mu \otimes \nu, T_a)$  has zero entropy.*

Remark that the condition 2. is satisfied as soon as the function  $a$  is integrable and has zero integral (cf [2] or [3]). Furthermore it is a simple consequence of Rokhlin Lemma and Baire Theorem that, as soon as the dynamical system  $(X, \mu, T)$  is aperiodic, there exist (a lot of) integrable functions  $a$ , with zero mean, satisfying the condition 1. (See for example the first theorem in [18]).

As usual, we write  $a^{(n)}(x) := \sum_{k=0}^{n-1} a(T^k x)$ .

**Lemma 5.5.** *For  $\mu$ -almost every  $x \in X$ , for all  $t \in \mathbb{Z}$ ,*

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=0}^{N-1} \mathbf{1}_{\{t\}} \left( a^{(n)}(x) \right) = 0 .$$

*Proof.* The existence of the limit is a consequence of the ergodic theorem. Let us fix  $\epsilon > 0$  and define

$$E(x) := \left\{ t \in \mathbb{Z} : \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=0}^{N-1} \mathbf{1}_{\{t\}} \left( a^{(n)}(x) \right) > \epsilon \right\}$$

This set is finite and, if it is not empty, then we have

$$\max(E(Tx)) = \max(E(x)) - a(x) .$$

Using condition 1. we obtain that, almost surely,  $E(x) = \emptyset$ . □

*Proof of Proposition 5.4.* For each  $x \in X$ , satisfying the property described in Lemma 5.5, we are going to define a transformation  $\varphi_x$  of  $\Omega$ . We construct, in a measurable way and by induction, an increasing sequence of nonnegative integers  $(N_k)_{k \geq 0}$  such that  $N_0 = 0$  and

$$\text{if } E_k := \left\{ a^{(n)}(x) : 0 \leq n \leq N_k \right\}, \quad \text{then } \frac{1}{N_{k+1}} \sum_{i=1}^{N_{k+1}} \mathbf{1}_{E_k} \left( a^{(i)}(x) \right) < \frac{1}{10} .$$

If  $\omega = (\omega(n))_{n \in \mathbb{Z}} \in \Omega$ , we pose

$$(\varphi_x(\omega))(n) = \begin{cases} \omega(n) & \text{if } n \leq 0, \\ 1 - \omega(n) & \text{if } n \in E_{k+1} \setminus E_k \text{ with } k \text{ even,} \\ \omega(n) & \text{if } n \in E_{k+1} \setminus E_k \text{ with } k \text{ odd.} \end{cases}$$

For any  $\omega \in \Omega$ , the sequence

$$\left( \frac{1}{N} \sum_{n=0}^{N-1} (-1)^{\omega(a^n(x)) + \varphi_x(\omega)(a^n(x))} \right)$$

does not converge.

Now we define a transformation  $\varphi$  of  $X \times \Omega$  by

$$\varphi(x, \omega) = (x, \varphi_x(\omega)) ,$$

and we consider the function  $f$  defined on  $X \times \Omega$  by  $f(x, \omega) = (-1)^{\omega(0)}$ .

The transformation  $\varphi$  preserves the product measure  $\mu \otimes \nu$  and the sequence

$$\left( \frac{1}{N} \sum_{n=0}^{N-1} f(T_a^n(x, \omega)) \cdot f(T_a^n(\varphi(x, \omega))) \right)$$

does not converge. Using Proposition 5.1, we conclude that  $T_a$  is not self-weakly disjoint.

Let us now show quickly why the recurrence hypothesis 2. implies that  $T_a$  as zero entropy as soon as  $T$  has.

Let  $A, B$  be measurable subsets of  $X$  and  $\Omega$  respectively. We suppose that  $T$  has zero entropy, and we choose  $A$  such that the partition  $\{A, A^c\}$  is a generator for the action of  $T$  on  $X$ . Such sets  $A$  form a dense class in the  $\sigma$ -algebra  $\mathcal{A}$ . We denote by  $\mathcal{C}$  the  $\sigma$ -algebra of subsets of  $X \times \Omega$  generated by  $(T_a^{-n}(\{A, A^c\} \otimes \{B, B^c\}), n > 0)$ . In order to prove that  $T_a$  has zero entropy, it is sufficient to show that the partition  $\{A, A^c\} \otimes \{B, B^c\}$  is  $\mathcal{C}$ -measurable. We have  $\mathcal{A} \otimes \{\emptyset, \Omega\} \subset \mathcal{C}$ . By the hypothesis 2., we can write  $A = \bigcup_{n>0} A_n \pmod{\mu}$ , where  $a^{(n)}(x) = 0$  for  $x \in A_n$ . For each  $n > 0$ , the event  $(x \in A_n \text{ and } T_a^n(x, y) \in X \times B)$  belongs to  $\mathcal{C}$ . But this event is equal to  $A_n \times B$ . We conclude that  $A \times B \in \mathcal{C}$ , which gives  $\{A, A^c\} \otimes \{B, B^c\} \subset \mathcal{C}$ .

**5.4. Rank one constructions.** We describe now a method of construction of rank one dynamical systems that are not self-weakly disjoint. This method is flexible enough to give us weakly mixing rigid examples, as well as strongly mixing examples.

**5.4.1. Reminder for the cutting and stacking construction of rank one dynamical systems.** Let us recall the general method of rank one system construction. At the first stage we consider an indexed family of  $h_1 \geq 2$  real disjoint intervals of the same length denoted  $B_1, TB_1, \dots, T^{h_1-1}B_1$ . Such a family is called a *tower of base  $B_1$  and height  $h_1$* . The intervals  $T^k B_1$  are the *levels* of the tower. At this stage the transformation  $T$  is defined on  $\bigcup_{0 \leq k \leq h_1-2} T^k B_1$  by the fact that it sends  $T^k B_1$  onto  $T^{k+1} B_1$  by translation. The transformation  $T$  is not defined on  $T^{h_1-1} B_1$ .

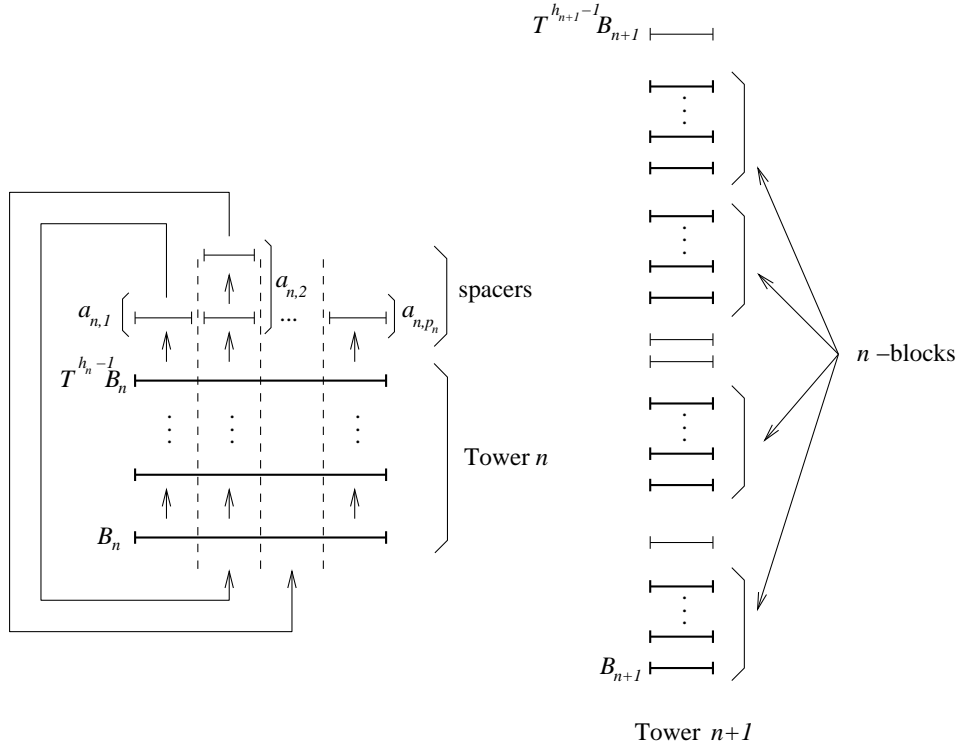
Stage  $n$  of the construction is given by a tower  $(B_n, TB_n, \dots, T^{h_n-1} B_n)$ , of base  $B_n$  and height  $h_n$ , called *Tower  $n$* .

Let us describe now how we go from Tower  $n$  to Tower  $n+1$ . This transition is parametrized by natural integers  $p_n \geq 2$  and  $a_{n,i} \geq 0$ ,  $1 \leq i \leq p_n$ . We cut the base  $B_n$  in  $p_n$  intervals  $I_{n,i}$ ,  $1 \leq i \leq p_n$ , of the same length. The base of Tower  $n+1$  is  $B_{n+1} := I_{n,1}$ . The levels  $T^k B_{n+1}$ ,  $1 \leq k \leq h_n-1$  are the subintervals of levels  $T^k B_n$  given by the definition of  $T$  at stage  $n$ . Then we consider  $a_{n,1}$  intervals  $S_1, \dots, S_{a_{n,1}}$  with the same length as  $B_{n+1}$ , pairwise disjoint and disjoint

from any of the intervals used before. These new intervals are called the *spacers*. For  $1 \leq j \leq a_{n,1}$ , we pose

$$T^{h_n+j-1}B_{n+1} := S_j.$$

Then we come back into  $B_n$  by posing  $T^{h_n+a_{n,1}}B_{n+1} := I_{n,2}$ . We repeat this procedure starting from  $I_{n,2}$ , adding this time  $a_{n,2}$  spacers before coming back onto  $I_{n,3}$ , and so on until  $T^{h_n-1}I_{n,p_n}$  above which we add  $a_{n,p_n}$  spacers.



Tower  $n+1$  so defined contains  $p_n$  slices of Tower  $n$ , called the  $n$ -blocks, and which are the towers  $I_{n,i}, TI_{n,i}, \dots, T^{h_n-1}I_{n,i}$ . Between these blocks,  $a_{n,1} + \dots + a_{n,p_n}$  spacers are inserted. The height is

$$h_{n+1} = p_n h_n + a_{n,1} + \dots + a_{n,p_n},$$

and the definition of the transformation  $T$ , on all the levels of Tower  $n+1$  but the last one, is compatible with the definition at the preceding stage.

Given the initial height  $h_1$  and parameters  $p_n$  and  $a_{n,i}$  ( $n \geq 1$ ,  $1 \leq i \leq p_n$ ) it is always possible to construct an infinite sequence of towers. Let us denote by  $X$  the union of all intervals which appear in the construction. Under the condition

$$(5.1) \quad \sum_{n \geq 1} \frac{a_{n,1} + \dots + a_{n,p_n}}{p_n h_n} < +\infty,$$

the Lebesgue measure of  $X$  is finite, and, changing if necessary the length of  $B_1$ , we can suppose that the measure of  $X$  is 1. In all the sequel, the condition (5.1) is supposed to be satisfied. The transformation  $T$  is almost everywhere defined on  $X$ , and it preserves the measure : we obtain what is called a *rank one dynamical system*. Such a system is always ergodic.

We call *transition  $n$*  the transition from Tower  $n$  to Tower  $n+1$ .

Let us give a few more useful definitions. If  $T^i B_n$  and  $T^j B_n$  are two levels of Tower  $n$ , where  $i$  and  $j$  are between 0 and  $h_n - 1$ , we call the *height difference in Tower  $n$*  between these levels the number

$$d_n(T^i B_n, T^j B_n) := j - i.$$

If  $x$  and  $y$  are two points belonging to the levels  $T^i B_n$  and  $T^j B_n$  respectively, we define similarly the height difference in Tower  $n$  between  $x$  and  $y$  by

$$d_n(x, y) := d_n(T^i B_n, T^j B_n) = j - i.$$

Finally, we note  $\mu_n$  the measure of the union of the levels of Tower  $n$ . The sequence  $(\mu_n)$  increases and goes to 1 as  $n \rightarrow \infty$ .

**5.4.2. Classical examples.** Using different choices for the parameters  $p_n$  and  $a_{n,i}$ , we obtain dynamical systems with various properties, going from discrete spectrum to strong mixing. Let us give now three classical examples of transitions. We use these examples in the sequel.

*The flat transition.* This transition is the simplest that we can imagine : all  $a_{n,i}$  are zero, we add no spacer. In the first and simplest example of rank one dynamical system, the Von Neumann-Kakutani transformation, all the transitions are flat and  $p_n = 2$  for all  $n$ .

It is not difficult to verify that if in the rank one construction there are flat transitions with  $p_n$ 's arbitrarily large, then the dynamical system is rigid.

*Chacon's transition.* In the construction of Chacon's transformation the transition  $n$  is described for all  $n$  by  $p_n = 3$ ,  $a_{n,1} = a_{n,3} = 0$  and  $a_{n,2} = 1$  : there is only one spacer and it is put on the middle column. We give the name of Chacon to this transition. Chacon's transformation is weakly mixing but not strongly mixing.

More generally, if in the rank one construction there are infinitely many Chacon's transitions, then the dynamical system is weakly mixing and not strongly mixing (see for example [7]).

*Staircase transition.* The transition  $n$  is called a *staircase transition* if  $a_{n,i} = i - 1$ , for  $1 \leq i \leq p_n$ . This transition is the key of the technics we want to describe.

Adams ([1]) has shown that if a rank one dynamical system is constructed with staircase transitions at each step, if  $\lim_{n \rightarrow \infty} p_n = +\infty$  and  $\lim_{n \rightarrow \infty} p_n/n^d = 0$  for some  $d > 0$ , then the system is strongly mixing.

**5.4.3. Construction of a rank one dynamical system which is not self-weakly disjoint.** We want to describe simultaneously the construction of a rank one system following the method described in 5.4.1, and a measure preserving transformation  $\varphi$  of  $X$ , such that, for almost all  $x$ , the sequence

$$m_N(x) := \frac{1}{N} \sum_{k=0}^{N-1} \mathbf{1}_{B_1}(T^k x) \mathbf{1}_{B_1}(T^k \varphi(x))$$

does not converge. By Proposition 5.1, such a system is not self-weakly disjoint. We obtain  $\varphi$  as the limit of a sequence  $(\varphi_n)$ , where, for each  $n$ ,  $\varphi_n$  is a permutation of the levels of Tower  $n$ . The permutations  $\varphi_n$  must satisfy the following compatibility condition : if  $E_{n+1}$  is a level of Tower  $n+1$  contained in the level  $E_n$  of Tower  $n$ , then  $\varphi_{n+1}(E_{n+1}) \subset \varphi_n(E_n)$ . Under this condition the limit  $\varphi$  is well defined and measure preserving.

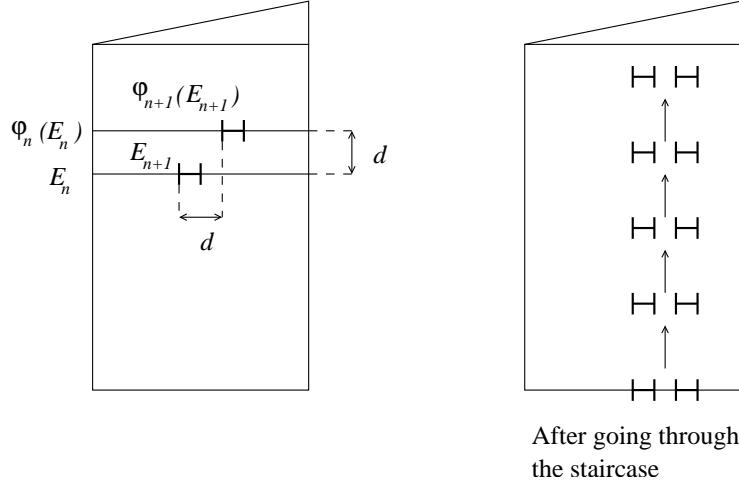
As for the transition from a tower to the next one, there exist several methods to construct  $\varphi_{n+1}$  starting from  $\varphi_n$ . We will describe three of these methods and their properties relatively to the averages  $m_N$ .

*Method 1 : the simplest.* If the permutation  $\varphi_n$  of the levels of Tower  $n$  is given, the simplest method to define  $\varphi_{n+1}$  compatible with  $\varphi_n$  is to let each of the  $n$ -blocks globally invariant by  $\varphi_{n+1}$ , and each of the spacers fixed. Let us denote

$$\gamma_n := \max_{E_n} \left| d_n(E_n, \varphi_n(E_n)) \right|.$$

If we define  $\varphi_{n+1}$  starting from  $\varphi_n$  and using method 1, we have  $\gamma_{n+1} = \gamma_n$ .

*Method 2 : To glue  $T^k x$  and  $T^k \varphi(x)$ .* This method can be used only if the transition  $n$  is of staircase type. Let  $E_n$  be a level of Tower  $n$ , and  $d := \left| d_n(E_n, \varphi_n(E_n)) \right|$ . Let  $E_{n+1}$  be a level of Tower  $n+1$  contained in  $E_n$ . For  $\varphi_{n+1}(E_{n+1})$ , we choose the piece of  $\varphi_n(E_n)$  which is shifted to  $d$  columns to the left (respectively to the right), if  $\varphi_n(E_n)$  stands below (respectively above)  $E_n$  in Tower  $n$ . If this shift is impossible because  $E_{n+1}$  is in the first  $d$  or in the last  $d$   $n$ -blocks, the shift is calculated modulo  $p_n$ . When  $E_{n+1}$  is a level of Tower  $n+1$  which is contained in none of the levels of Tower  $n$  (that is to say when  $E_{n+1}$  is a spacer), we pose  $\varphi_{n+1}(E_{n+1}) := E_{n+1}$ .



Let  $x$  be a point of level  $E_{n+1}$  which is not a spacer and which is neither in the first  $\gamma_n$   $n$ -blocks, nor in the last  $\gamma_n$   $n$ -blocks. After going through the staircase, the points  $T^k x$  and  $T^k \varphi(x)$  come back simultaneously in  $B_n$ , and climb together Tower  $n$ . For the  $h_n$  indices between  $k$  and  $2h_n + p_n - 1$  which correspond to this first complete climbing, we have  $\mathbf{1}_{B_1}(T^k x) = \mathbf{1}_{B_1}(T^k \varphi(x))$ . Denoting by  $b_n$  the number of indices  $k \in \{0, \dots, h_n - 1\}$  such that  $T^k B_n \subset B_1$ , we have

$$(5.2) \quad m_{2h_n+p_n}(x) \geq \frac{h_n}{2h_n + p_n} \frac{b_n}{h_n}.$$

We remark that the measure of the set of points  $x$  for which this inequality is not satisfied is bounded by  $1 - \mu_n + 2\gamma_n/p_n$ . We remark also that

$$b_n/h_n = \mu(B_1)/\mu_n \longrightarrow \mu(B_1),$$

when  $n \rightarrow \infty$  and that

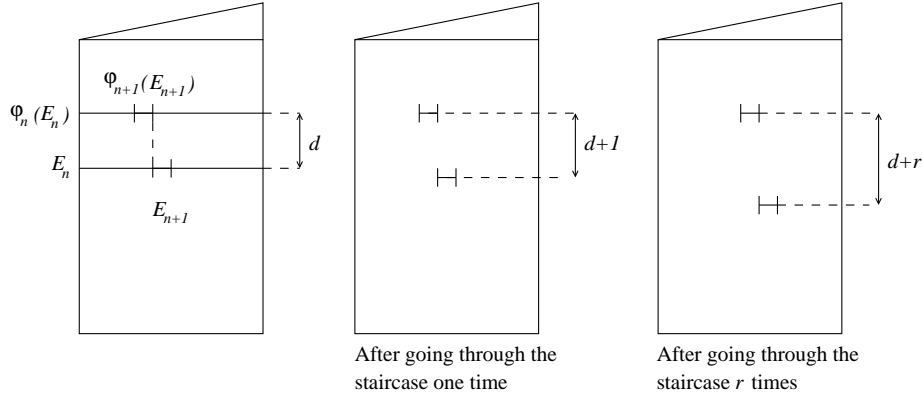
$$\frac{p_n - 1}{h_n} \leq 1 - \mu_n.$$

Hence the right term in (5.2) is close to  $\mu(B_1)/2$  when  $\mu_n$  is close to 1.



*Method 3 : mixing.* As in the preceding method, here we consider only staircase transitions. If  $E_{n+1}$  is a level of Tower  $n+1$  contained in the level  $E_n$  of Tower  $n$ , we choose  $\varphi_{n+1}(E_{n+1})$  in  $\varphi_n(E_n)$  by shifting to one column on the left. As in method 2, we calculate the shift modulo  $p_n$  when  $E_{n+1}$  is in the first  $n$ -block, and we let the spacers fixed under  $\varphi_{n+1}$ .

Let  $r$  be a positive integer, and  $x \in E_{n+1} \subset E_n$ , where  $E_n$  and  $E_{n+1}$  are levels of Tower  $n$  and Tower  $n+1$  respectively. In particular,  $x$  is not in a spacer of the preceding step. We suppose also that  $x$  does not belong to the first or the last  $(r+1)$   $n$ -blocks. The measure of the set of points excluded by these conditions is bounded by  $1 - \mu_n + (r+2)/p_n$ . Let  $d := d_n(E_n, \varphi_n(E_n))$ . After going through the staircase once, the height difference between  $T^k x$  and  $T^k \varphi(x)$  in Tower  $n$  becomes  $d+1$ ; after going through the staircase  $j$  times, it becomes  $d+j$  ( $1 \leq j \leq r$ ).



When  $k$  goes from 0 to  $(r+1)(h_n + p_n) - 1$ , the point  $T^k x$  climbs (at least)  $r$  times Tower  $n$ . Denote by  $G$  the union of all the levels of Tower  $n$ , except the first  $\gamma_n$  ones and the last  $r + \gamma_n$  ones, and denote  $J$  the set of the  $r(h_n - r - 2\gamma_n)$  indices  $k \in \{0, \dots, (r+1)(h_n + p_n) - 1\}$  which correspond to the times of the first  $r$  climbings of  $T^k x$  in  $G$ . We have

$$(5.3) \quad \frac{1}{|J|} \sum_{k \in J} \mathbf{1}_{B_1}(T^k x) \mathbf{1}_{B_1}(T^k \varphi(x)) = \frac{r}{|J| \mu(B_n)} \int_G \mathbf{1}_{B_1}(y) \left( \frac{1}{r} \sum_{j=1}^r \mathbf{1}_{B_1}(T^{d+j} y) \right) d\mu(y).$$

Since

$$\frac{|J|}{(r+1)(h_n + p_n)} = \frac{r}{r+1} \frac{h_n - r - 2\gamma_n}{h_n + p_n},$$

the left term in (5.3) is a good approximation of  $m_{(r+1)(h_n + p_n)}(x)$  if the number  $r$  is big enough and if  $h_n$  is big enough with respect to  $(p_n + \gamma_n + r)$ . Adding to these conditions the fact that  $\mu_n$  is close to 1, the right term of (5.3) is close to

$$I_{d,r} := \int_X \mathbf{1}_{B_1}(x) \left( \frac{1}{r} \sum_{j=1}^r \mathbf{1}_{B_1}(T^{d+j} x) \right) d\mu(x).$$

By the ergodicity of  $T$ , the sequence  $\frac{1}{r} \sum_{j=1}^r \mathbf{1}_{B_1}(T^j x)$  goes to  $\mu(B_1)$  in probability, and we see that, if  $r$  is big enough (independently of  $d$ ),  $I_{d,r}$  is close to  $\mu(B_1)^2$ .

5.4.4. *Results.* These three methods for the construction of permutations  $\varphi_n$  give us sufficient conditions for a rank one dynamical system to be not self-weakly disjoint.

**Theorem 5.6.** *Let  $T$  be a rank one dynamical system constructed as in 5.4.1. If, for all integer  $M$ , there exists an integer  $n$  such that the transition  $n$  is of staircase type with a number of steps  $p_n > M$ , then  $T$  is not self-weakly disjoint.*

*Proof.* Since there is an infinity of staircase transitions in the construction of  $T$ , it is possible to construct a sequence of permutations  $(\varphi_n)$  using infinitely often methods 2 and 3. Let us describe more precisely the inductive construction of  $(\varphi_n)$ . We pose  $n_0 := 0$  and we choose  $\varphi_1$  arbitrarily. Suppose that  $\varphi_1, \dots, \varphi_{n_k+1}$  are already constructed. By hypothesis, there exists  $n'_k \geq n_k + 1$  such that the transition  $n'_k$  is of staircase type, with  $p_{n'_k} > 2^k \gamma_{n_k+1}$  and  $\mu_{n'_k} > 1 - 2^{-k}$ . We construct  $\varphi_{n_k+2}, \dots, \varphi_{n'_k}$  using always method 1 to keep constant the value of  $\gamma_n$ , and then we construct  $\varphi_{n'_k+1}$  by method 2.

Afterward we consider an integer  $r_k > 2^k$  such that, for all  $d$ ,  $|I_{r_k, d} - \mu(B_1)^2| < 2^{-k}$  (this is satisfied by any large enough  $r_k$ ). Let  $n_{k+1}$  be the first integer larger than  $n'_k + 1$  such that the transition  $n_{k+1}$  is of staircase type, with  $p_{n_{k+1}} > 2^k r_k$  and  $h_{n_{k+1}} > 2^k(\gamma_{n'_k+1} + r_k + p_{n_{k+1}})$ . We construct  $\varphi_{n'_k+2}, \dots, \varphi_{n_{k+1}}$  by method 1, then  $\varphi_{n_{k+1}+1}$  by method 3.

If we repeat this procedure, the transformation  $\varphi$  defined as the limit of the sequence  $(\varphi_n)$  is such that, for  $\mu$ -almost every  $x$ ,

$$\limsup_{N \rightarrow \infty} m_N(x) \geq \frac{1}{2} \mu(B_1),$$

and

$$\liminf_{N \rightarrow \infty} m_N(x) \leq \mu(B_1)^2.$$

Since  $\mu(B_1) < 1/2$ , this proves that  $T$  is not self-weakly disjoint.  $\square$

**Corollary 5.7.** *There exist strongly mixing rank one dynamical systems which are not self-weakly disjoint. There exist also weakly mixing and rigid rank 1 dynamical systems which are not self-weakly disjoint.*

*Proof.* The mixing rank one systems described in [1] satisfy the hypothesis of Theorem 5.6.

We can also construct rank one systems with an infinity of flat transitions, giving rigidity, an infinity of Chacon's transitions, giving weak mixing, and an infinity of staircase transitions, with  $p_n$  going to  $+\infty$ , so that Theorem 5.6 applies.  $\square$

**Corollary 5.8.** *There exist dynamical systems with minimal self-joinings which are not self-weakly disjoint.*

Indeed, any mixing rank one system has minimal self-joinings (see [13]).  $\square$

## REFERENCES

- [1] T.M. Adams, *Smorodinsky's Conjecture on rank-one mixing*, Proc. Amer. Math. Soc., **126**, p.739-744, 1998.
- [2] G. Atkinson, *Recurrence of cocycles and random walks*, J. London Math. Soc., **13**, p. 486-488, 1976.
- [3] F.M. Dekking, *On transience and recurrence of generalized random walks*, Z. Wahrscheinlichkeitstheorie verw. Gebiete, **61**, p. 459-465, 1982.

- [4] A. del Junco and M. Keane, *On generic points in the Cartesian square of Chacón's transformation*, Ergodic Theory and Dynamical Systems, **5**, p. 59-69, 1985.
- [5] A. del Junco, M. Rahe and L. Swanson, *Chacon's automorphism has minimal self joinings*, J. Anal. Math., **37**, p. 276-284, 1980.
- [6] A. del Junco and D.J. Rudolph, *On ergodic actions whose self-joinings are graphs*, Ergodic Theory and Dynamical Systems, **7**, p. 531-557, 1987.
- [7] N.A. Friedman, *Replication and Stacking in Ergodic Theory*, Amer. Math. Monthly, **99**, p.31-41, 1992.
- [8] H. Furstenberg, *Disjointness in ergodic theory, minimal sets, and a problem in diophantine approximation*, Math. Systems Theory, **1**, p. 1-49, 1967.
- [9] H. Furstenberg, *Recurrence in Ergodic Theory and Combinatorial Number Theory*, Princeton UP, 1981.
- [10] E. Ghys, *Dynamique des flots unipotents sur les espaces homogènes*, Séminaire Bourbaki, No 747. Astérisque (SMF) **206**, 1991.
- [11] P.R. Halmos, *Lectures on Ergodic Theory*, Chelsea Pub.Co, New-York, 1956.
- [12] A.S. Kechris, *Classical Descriptive Set Theory*, Springer-Verlag, 1995.
- [13] J.L. King, *Joining-rank and the structure of finite-rank mixing transformations*, J. Analyse Math., **51**, p.182-227, 1988.
- [14] M. Lemańczyk, F. Parreau and J.P. Thouvenot, *Gaussian automorphisms whose ergodic self-joinings are Gaussian*, Fundamenta Mathematicae, **164**, p. 253-293, 2000.
- [15] E. Lesigne, *Spectre quasi-discret et théorème ergodique de Wiener-Wintner pour les polynômes*, Ergodic Theory and Dynamical Systems, **5**, p. 767-784, 1993.
- [16] E. Lesigne, C. Mauduit and B. Mossé, *Le théorème ergodique le long d'une suite  $q$ -multiplicative*, Compositio mathematica, **93**, p. 49-79, 1994.
- [17] E. Lesigne, A. Quas and M. Wierdl, *Generic points in the Cartesian powers of the Morse dynamical system*, preprint.
- [18] E. Lesigne and D. Volny, *Large deviations for generic stationary processes*, Colloquium Mathematicum **84/85**, p. 75-82, 2000.
- [19] C. Mauduit, *Substitutions et ensembles normaux*, Habilitation à diriger des recherches, Univ. Aix-Marseille II, 1989.
- [20] M. Ratner, *Invariant measures for unipotent translations on homogeneous spaces*, Proc. Nat. Acad. Sci. USA **87**, 4309-4311, 1990.
- [21] B. Rittaud, *Convergence ponctuelle de moyennes ergodiques non conventionnelles et distribution asymptotique de suites oscillantes*, Thèse de Doctorat, Université de Tours, 1999.
- [22] A. Rothstein, *Versik processes: First steps*, Israel J. Math. **36**, p. 205-224, 1980.
- [23] D.J. Rudolph, *An example of measure preserving map with minimal self-joinings, and applications*, J. Anal. Math., **35**, p. 97-122, 1979.
- [24] J.P. Thouvenot, *Some properties and applications of joinings in Ergodic Theory*, Ergodic Theory and its Connections with Harmonic Analysis, London Math. Soc. Lecture Note Series **205**, p. 207-235, 1995.

(Emmanuel Lesigne) LABORATOIRE DE MATHÉMATIQUES ET PHYSIQUE THÉORIQUE, UMR CNRS 6083, UNIVERSITÉ FRANÇOIS RABELAIS, PARC DE GRANDMONT, 37200 TOURS, FRANCE  
*E-mail address:* `lesigne@univ-tours.fr`

(Benoît Rittaud) LABORATOIRE D'ANALYSE, GÉOMÉTRIE ET APPLICATIONS, UMR CNRS 7539, INSTITUT GALILÉE, UNIVERSITÉ PARIS 13, 99, AV. J.B. CLÉMENT, 93430 VILLETANEUSE, FRANCE  
*E-mail address:* `rittaud@math.univ-paris13.fr`

(Thierry de la Rue) LABORATOIRE DE MATHÉMATIQUES RAPHAËL SALEM, UMR CNRS 6085, UNIVERSITÉ DE ROUEN, 76821 MONT SAINT-AIGNAN, FRANCE  
*E-mail address:* `thierry.delarue@univ-rouen.fr`